

[Subscribe to KDnuggets](#)

[Submit a blog](#)
[Win a Reward!](#)

[Blog](#) | [Opinions](#) | [Tutorials](#) | [Top stories](#) | [Courses](#) | [Datasets](#) | [Education: Online](#) | [Certificates](#) | [Events / Meetings](#) | [Jobs](#) | [Software](#) | [Webinars](#)

We asked
300 leaders...

What's holding your
data science team back?

Get t

The 2021 Domino Data Lab Maturity Index

[Topics: AI](#) | [Data Science](#) | [Data Visualization](#) | [Deep Learning](#) | [Machine Learning](#) | [NLP](#) | [Python](#)

[KDnuggets Home](#) » [News](#) » [2018](#) » [Aug](#) » [Tutorials, Overviews](#) » [Comparison of the Most Useful Text Process](#)

Comparison of the Most Useful Text Processing APIs



[<= Previous post](#)

[Next post =>](#)

Like 61

Share 61

Tweet

Share

Share

39

Tags: [NLP](#), [Text Analytics](#), [Text Mining](#)

There is a need to compare different APIs to understand key pros and cons they have and when it is better to use one API instead of the other. Let us proceed with the comparison.



[Anaconda Nucleus](#)
[Exclusive tutorials](#)
[Interactive notebooks](#)

[comments](#)

By [ActiveWizards](#)



Nowadays, text processing is developing rapidly, and several big companies provide their products which help to deal successfully with diverse text processing tasks. In case you need to do some text processing there are 2 options available. The first one is to develop the entire system on your own from scratch. This way proves to be very time and resource consuming. On the other hand, you can use the already accessible solutions developed by well-known companies. This option is usually faster and simpler. No specific knowledge or experience in the natural language processing is required. It would suffice to understand the fundamentals of the text processing. At the same time, if you need something exclusive, it is better to implement own solution rather than to apply one of the above mentioned.

Anyway, APIs for text processing are very popular and useful. So, there is a need to compare different APIs to understand key pros and cons they have and when it is better to use one API instead of the other. Let us proceed with the comparison.

Working with text processing, the data analyst faces the following tasks:

- Keyphrase extraction;
- Sentiment analysis;
- Text analysis;
- Entity recognition;
- Translation;
- Language detection;
- Topic modeling.

There are several high-level APIs which may be used to perform these tasks. Among them:

- [Amazon Comprehend](#);
- [IBM Watson Natural Language Understanding](#);
- [Microsoft Azure \(Text analytics API\)](#);
- [Google Cloud Natural Language](#);
- [Microsoft Azure \(Linguistic Analysis API\) - beta](#);
- [Google Translate API](#);

Enhance your AI Experimentation
 Run hyperparameter optimization in parallel with a few lines of code

SIGOPT

Run hyperparameter optimization parallel with a few lines of code

Latest News

- [Data Science Web nugget Roundup Jan 29: Calculating De...](#)

Get The Latest News!

- [Classification with HuggingFace](#)
- [Archives of the KDnuggets, Leading newsletter in Data Science and Machine Learning](#)
- [What to Expect From Your Career Path as a Data Scientist](#)

Email:

[Sign Up](#)

- By subscribing you accept KDnuggets [Privacy Policy](#)
- [R vs Python \(Again\): A Human Factor Perspective](#)

Top Stories Last Week

1. [The High Paying Side Hustles for Data Scientists](#)



2. [Is Data Science a Dying Career](#)
3. [Top Programming Languages and Their Uses](#)
4. [Models Are Rarely Deployed: A Industry-wide Failure in Machine Learning Leadership](#)
5. [How to Process a DataFrame with Millions of Rows in Seconds](#)

More Recent Stories

[Subscribe to KDnuggets News](#)



Subscribe to
KDnuggets



Submit a blog
Win a Reward!

[Blog](#) | [Opinions](#) | [Tutorials](#) | [Top stories](#) | [Courses](#) | [Datasets](#) | [Education: Online](#) | [Certificates](#) | [Events / Meetings](#) | [Jobs](#) | [Software](#) | [Webinars](#)

Amazon Comprehend

[Amazon Comprehend](#) is an element of the Amazon Web Services infrastructure. It provides high-level API for such text-processing tasks as language detection, sentiment analysis, topic modeling, keyphrase extraction, and entity recognition.

In case of the **language detection**, the Amazon Comprehend is capable of analyzing the input and then returning the dominant language and the confidence score. At the moment the tool supports more than 100 languages for this task.

Keyphrase extraction increases the understanding of the text you analyze. Amazon Comprehend returns the list of all detected key phrases with the confidence scores for each of them.

Entity recognition is the process of splitting the text into entities: people, organizations, locations, dates, etc. Thus, for each entity, the Amazon Comprehend will define a group to which it belongs. Like in the previous tasks, the confidence scores are also returned.

Sentiment analysis in the Amazon Comprehend allows detecting the general mood of the text. It supports 4 different sentiments: positive, negative, neutral, and mixed. The confidence scores are returned for each sentiment separately.

The outcomes of the **topic modeling** in Amazon Comprehend may be presented in 2 variants. According to the first variant, it will show the topic group for each document, the main keyword for this topic group, and the confidence score. According to the second variant, the topic group for each document and the proportion of the main keyword in the document are shown. A negative feature here is that if you want to perform topic modeling, you should have all your documents stored in Amazon S3.

Free Tier program is available up to 12 months. Here you pay just for those things you are using and only in the amounts required. Thus, Amazon Comprehend provides great flexibility in pricing. One of the downsides of this platform is that it supports only English and Spanish languages.

Nevertheless, you can use Amazon Translate (one more cloud service by the AWS) to translate from the source language into English or Spanish and then analyze the text using Amazon Comprehend.

IBM Watson Natural Language Understanding

IBM Watson is the platform which provides Artificial Intelligence related services for business. [Natural Language Understanding](#) is a part of this infrastructure. It allows solving a wide variety of tasks in text processing such as entity recognition, sentiment analysis, keyphrase extraction, topic modeling, and text analysis. Language detection is supported as a side feature (it returns the source language when you perform any task). There are some peculiarities in the solutions provided by Watson Natural Language Understanding. Let's explore them more carefully.

Entity recognition in Natural Language Understanding allows detecting people, dates, places, organizations, geopolitical entities, etc. It not only returns the confidence score but also returns the subtype of the entity. For example, when it deals with famous people, it does not only say that this is a person, but it also recognizes that this person is an actor, award winner, and so on. Also, it can even return a link to corresponding DBpedia resource.

Sentiment analysis with the help of Natural Language Understanding also provides additional helpful features. It can not only define whether the sentiment is positive or negative but can also return sentiment separately for each detected entity, keyword phrase or even predefined target piece of the sentence. The service is also capable of the emotion recognition (anger, disgust, fear, joy, or sadness) for the whole text as well as for separate detected keywords, entities or user-specified target phrases. Confidence scores are returned as well.

Keywords extraction here is pretty simple and standard. It returns keywords along with the confidence scores for each detected keyword.

Topic modeling is provided by returning up to 5 levels hierarchy of the text categories. Also, the system can detect more than 1 chain of categories. For example, if you feed the text about the Titanic, it can return one chain of categories concerning ships, and another chain concerning shipwrecks.

Text analysis in Watson Natural Language Understanding consists of three major parts: finding relations in the text, analyzing the semantic roles and extracting the text's metadata. The system can determine the relations between the entities. The analysis of the semantic roles allows detecting



TensorFlow for Computer Vision
Transfer Learning Made

Learn Machine Learning 4X Faster
by Participating in Competitions

Get The Latest News!

Stack on a Budget



Get KDnuggets,
leading newlett
on AI, Data Scien
and Machine
Learning

Top April Stories: The Most In-
Demand Skills for Data Scientists

Sign Up

Top November Stories: Top Python
Libraries for Data Science, Data...

Top January Stories: How to land a
Data Scientist job at your dream...

Top November Stories: How to
Speed up Pandas by 4x with one
line of code

Top July Stories: The Death of Big
Data and the Emergence of the...

Top August Stories: Know What
Employers are Expecting for a
Data...

Subscribe to KDnuggets News



Subscribe to
KDnuggets



Submit a blog
Win a Reward!

[Blog](#) | [Opinions](#) | [Tutorials](#) | [Top stories](#) | [Courses](#) | [Datasets](#) | [Education: Online](#) | [Certificates](#) | [Events / Meetings](#) | [Jobs](#) | [Software](#) | [Webinars](#)

Based on our experience, IBM Watson Natural Language Understanding has good performance. Despite there was a need to analyze and filter the outcomes of the service in several tasks, in general, the platform operates very well.

Microsoft Azure (Text Analytics API)

Text Analytics API is an element of the Microsoft Azure infrastructure. It proposes a relatively poor set of solutions. It supports only **language detection**, **keyphrase extraction**, and **sentiment analysis**. There is also a limited number of positive results each feature can return. For instance, there are no confidence scores for keyphrase extraction and sentiment may be either negative or positive. However, it supports a vast number of languages. In addition, the service produces perfect results within a short period of time.

There are a free tier and five options for a chargeable usage. The levels differ from each other by the number of maximum allowed transactions per year. Also, there may be some variations in prices across different regions (see this [page](#) for more details).

Microsoft Azure (Linguistic Analysis API Preview)

Linguistic Analytics API is a part of the Microsoft Azure platform too.

This product attempts to extend the NLP functionality of the previous service (Text Analytics API), which has limited functionality. The Linguistic Analysis API proposes such solutions as sentences separation, tokenization, Part-of-Speech tagging, relations in the sentences analysis.

Sentence separation is the first step in the analysis. Linguistic Analysis API splits the text into sentences and then tokenizes the words in them.

POS-tagging allows determining what word to what part of speech belongs. The tool can return input sentences with the POS-tag near each word.

In our opinion, the most interesting thing here is the ability of the Linguistic Analysis API to build the schema of the connections between words in the sentences. Out of this diagram, you can understand how and which words are connected with each other.

As the platform is on the stage of testing, there is no pricing for it yet. We suppose that the prices will also depend on the number of transactions within the system.

Google Cloud Natural Language

Google Cloud Natural Language is a part of the Google Cloud infrastructure. It enables you to perform such tasks as sentiment analysis, entity recognition, topic modeling, and text analysis.

Sentiment analysis in this system allows detecting the sentiment of the sentences, or for the entire document and for each word in the sentences. The API returns the score in the range from -1 to 1, where -1 means very negative sentiment and 1 means very positive. In addition, the system detects the magnitude of the sentiment (the strength) which can vary from 0 to infinity.

Entity recognition detects types of entities in the text and returns the importance of each word to the entire text document (in the range of 0 to 1).

Topic modeling allows building several chains of categories of the text and the confidence score for each chain.

Text analysis seems to be quite effectual in the Google Cloud Natural Language. It can find the relations in the sentences, detect POS-tags, lemmas of the words. It also provides morphological analysis (number, proper tense, gender, person, a case of the word, etc.). The relations between the words may be represented in the form of the dependency trees.

Google Cloud Translation API

This instrument provides both **machine translation** and **language detection**. It is simple to integrate **Google Cloud Translation API** with your service. It uses deep neural networks for translation. These networks improve continuously. There is no need to extract and send a text to the API, you can send



Run hyperparameter optimization parallel with a few lines of code

Get The Latest News!



Get KDnuggets,
leading newslett
on AI, Data Scien
and Machine
Learning

Email:

[Sign Up](#)

By subscribing you accept KDnuggets

[Privacy Policy](#)

[Subscribe to KDnuggets News](#)



Subscribe to
KDnuggets



Submit a blog
Win a Reward!

[Blog](#) | [Opinions](#) | [Tutorials](#) | [Top stories](#) | [Courses](#) | [Datasets](#) | [Education: Online](#) | [Certificates](#) | [Events / Meetings](#) | [Jobs](#) | [Software](#) | [Webinars](#)

IBM Watson Translator

IBM Watson infrastructure contains [IBM Watson Translator](#) for **neural machine translation**. The main downside of this tool is the limited number of the supported languages. The main benefit is that it allows customization. For instance, if your texts include specific terminology the IBM Watson Translator can adapt to this terminology to work with it properly. It also includes some pre-trained language models like News, Conversation, etc. However, the models are available for an even smaller number of languages.

The pricing is flexible. You can choose out of 4 available levels (Lite, Standard, Advanced, and Premium). Each new level provides more functionality and allows processing more characters, so correspondingly the price is higher.

Amazon Translate

[Amazon Translate](#) is an element of the Amazon Web Services infrastructure. It allows fast translating in the real-time as well as by feeding the entire documents (batch). The tool is not well developed yet, so it supports only 12 languages. It isn't available for customization and doesn't use language models. However, developers state, the service will be improved in the near future, for example, 6 more languages will be added.

The Free Tier is available during the first 12 months. Then, you will have to pay 15 USD per 1 million processed characters.

Microsoft Azure Translator Text API

[Microsoft Azure Translator Text API](#) uses deep neural networks to improve the quality of the **machine translation**. This method takes into account not only the neighboring words but also the context of the entire sentence. The documentation is quite simple and clear. The tool also allows users to customize translations.

Microsoft suggests 5 different pricing plans. Processing less than 1 million characters per month is free of charge. If you choose an S1 plan, you will have to pay 10 USD per million characters. S2, S3, and S4 plans have fixed prices and the maximum amount of characters to process.

We have compiled a table comparing the main characteristics of these APIs for each separate task. Hopefully, it will help you to choose the API that is compatible with your needs.



Run hyperparameter optimization parallel with a few lines of code

Get The Latest News!



Get KDnuggets,
leading newslett
on AI, Data Scien
and Machine
Learning

Email:

By subscribing you accept KDnuggets

[Privacy Policy](#)

[Subscribe to KDnuggets News](#)



[Subscribe to KDnuggets](#)

[Submit a blog](#)
[Win a Reward!](#)

[Blog](#) | [Opinions](#) | [Tutorials](#) | [Top stories](#) | [Courses](#) | [Datasets](#) | [Education: Online](#) | [Certificates](#) | [Events / Meetings](#) | [Jobs](#) | [Software](#) | [Webinars](#)

	Language detection		Confidence scores returned		Performance
	Keyphrase extraction	Confidence scores returned	English, Spanish	GOOD	
Amazon Comprehend	Entity recognition	Confidence scores returned	English, Spanish	GOOD	FAST
	Sentiment analysis	Confidence scores returned, 4 different sentiments (positive, negative, neutral, mixed)	English, Spanish	GOOD	
	Topic modelling	2 views on the topics, documents should be stored in Amazon S3	English, Spanish	GOOD	
IBM Watson Natural Language Understanding	Language detection	Available as a side-feature	86	PERFECT	VERY FAST 2100 symbols less than 2 seconds
	Keyphrase extraction	Confidence scores returned	13	PERFECT	
	Entity recognition	Subtypes of entities, confidence scores, DBpedia resources returned	13	PERFECT GOOD	
	Sentiment analysis	Sentiments for piece of sentence, emotions, confidence scores returned	13	PERFECT	
	Topic modelling	5 levels hierarchy of categories, several chains of categories	13	PERFECT GOOD	
	Text analysis	Semantic roles, relations in the sentence, extracting metadata	13	GOOD INTERMEDIATE	
Microsoft Azure (Text Analytics API)	Language detection	Not a main feature	120	PERFECT	VERY FAST 2100 symbols about 2 seconds
	Keyphrase extraction	Without confidence scores	16	PERFECT	
	Sentiment analysis	Positive (100%) or negative (0%)	15	PERFECT	
Microsoft Azure (Linguistic Analysis API)	Text analysis	POS-tagging, tokenization, sentence separation, relations in the sentence (in the form of the diagram)	Unknown	GOOD	FAST
Google Cloud Natural Language	Entity recognition	Importances and Wikipedia articles	10	GOOD INTERMEDIATE	FAST
	Sentiment analysis	Document level as well as sentiment by sentences and entities. Magnitude of the sentiment	10	GOOD	
	Topic modelling	Confidence scores. Several topic chains	10	PERFECT	
	Text analysis	POS-tagging, relations in the sentences, lemma and morphology analysis	10	PERFECT	
Google Cloud Translation API	Translation and Language detection	Supports rendering HTML documents	100+	GOOD	VERY FAST 2100 symbols about 2 seconds
IBM Watson Translator	Translation	Supports customization by users	21	GOOD	FAST
Amazon Translate	Translation	Both batch and real-time translations	12 + 6 languages later	GOOD	FAST
Microsoft Azure Translator Text API	Translation and Language detection	Supports building customized translation systems	60+	GOOD	FAST

Created by [ActiveWizards](#)

Updated: August 2018

Conclusion

In this article, we highlighted and analyzed key characteristics of different APIs developed for text processing. Although all the APIs we mentioned pursue almost the same goals, they tend to have as many distinctive features as those in common. In order to select the best variant suiting your needs, you have to know for sure about the number of different languages you are going to work with. Another important issue to clarify is whether you are going to process the source text or you are looking for a helpful translation tool.

The majority of these APIs provide fast performance, and all of them are subjects for future improvements by the developers. Therefore, you always have some space for the development of your services. Finally, you should take into consideration the quality of the output along with the period of time within which it may be received.

These APIs have proved to be extremely useful, based on our experience, so we attempted to list all their key features in a comprehensible manner for you. However, there are many other APIs, which deserve to be looked at, so share your favorites in the comment section.

Thank you for your attention!

Enhance your AI Experimentation

Run hyperparameter optimization in parallel with a few lines of code

SIGOPT

Run hyperparameter optimization parallel with a few lines of code

Get The Latest News!



Get KDnuggets, leading newsletter on AI, Data Science and Machine Learning

Email:

[Sign Up](#)

By subscribing you accept KDnuggets

[Privacy Policy](#)

[Subscribe to KDnuggets News](#)



Subscribe to
KDnuggets



Submit a blog
Win a Reward!

[Blog](#) | [Opinions](#) | [Tutorials](#) | [Top stories](#) | [Courses](#) | [Datasets](#) | [Education: Online](#) | [Certificates](#) | [Events / Meetings](#) | [Jobs](#) | [Software](#) | [Webinars](#)

[Original](#). Reposted with permission.

Related:

- [Top 20 Python Libraries for Data Science in 2018](#)
- [Top 20 R Libraries for Data Science in 2018](#)
- [Top 15 Scala Libraries for Data Science in 2018](#)

0 Comments

KDnuggets

[Disqus' Privacy Policy](#)

Login

Favorite

Tweet

Share

Sort by Best



LOG IN WITH

OR SIGN UP WITH DISQUS

Be the first to comment.

Subscribe

Add Disqus to your site

Do Not Sell My Data

More On This Topic

- [Top October Stories: How to Become a \(Good\) Data Scientist;...](#)
- [Top May Stories: The Best NLP with Deep Learning Course is Free](#)
- [Top June Stories: How Much Math do you need in Data Science? Easy...](#)
- [Skills to Build for Data Engineering](#)
- [Top October Stories: Data Science Minimum: 10 Essential Skills You...](#)
- [Comparing Decision Tree Algorithms: Random Forest® vs. XGBoost](#)

[<= Previous post](#)

[Next post =>](#)

Top Stories Past 30 Days

1. [Is Data Science a Dying Career?](#)
2. [3 Tools to Track and Visualize the Execution of Your Python Code](#)
3. [Why Do Machine Learning Models Die In Silence?](#)
4. [6 Predictive Models Every Beginner Data Scientist Should Master](#)
5. [SQL Interview Questions for Experienced Professionals](#)
6. [Why are More Developers Using Python for Their Machine Learning Projects?](#)
7. [Top Five SQL Window Functions You Should Know For Data Science Interviews](#)
8. [What Makes Python An Ideal Programming Language For Startups](#)
9. [A Deep Look Into 13 Data Scientist Roles and Their Responsibilities](#)
10. [Top Programming Languages and Their Uses](#)

[KDnuggets Home](#) » [News](#) » [2018](#) » [Aug](#) » [Tutorials, Overviews](#) » Comparison of the Most Useful Text Processing APIs

© 2022 KDnuggets. | [About KDnuggets](#) | [Contact](#) | [Privacy policy](#) | [Terms of Service](#)



Run hyperparameter optimization parallel with a few lines of code

Get The Latest News!



Get KDnuggets, leading newslett on AI, Data Scien and Machine Learning

Email:

[Sign Up](#)

By subscribing you accept KDnugge

[Privacy Policy](#)

[Subscribe to KDnuggets News](#)

